

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES**

In re Application of:)	
)	
Andrei Z. Broder, et al.)	
)	
Serial No.: 10/057,331)	Group Art Unit: 2176
)	
Filed: January 24, 2002)	Examiner: Nathan Hillery
)	
For: METHOD FOR RANKING)	
WEB PAGE SEARCH RESULTS)	

Mail Stop: Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

In support of the Notice of Appeal filed on October 11, 2006 (received October 16, 2006), and pursuant to 37 C.F.R. § 41.37, Appellants present this appeal brief in the above-captioned application.

This is an appeal to the Board of Patent Appeals and Interferences from the Examiner's final rejection of claims 1-12 and 14-16 in the Final Office Action dated July 11, 2006. The appealed claims are set forth in the attached Claims Appendix.

1. Real Party in Interest

This application was originally assigned to Alta Vista, which has since been acquired by Yahoo! Inc., the real party in interest.

2. Related Appeals and Interferences

There are no other appeals or interferences which would directly affect, be directly affected by, or have a bearing on the instant appeal.

3. Status of Claims

Claims 1-12 and 14-16 have been rejected in the final Office Action. Claim 15 has been previously cancelled. The final rejection of claims 1-12 and 14-16 is being appealed.

4. Status of Amendments

All amendments submitted by the Appellants have been entered.

5. Summary of Claimed Subject Matter

Independent claim 1 recites a method for including a document in an index in a hyperlinked environment. (For example, see page 4, lines 24-28). The method includes receiving a document to be processed and locating a set of documents that include hyperlinks to the document. (For example, see page 6, lines 19-23). The method includes retrieving anchor text associated with at least one of the hyperlinks and parsing the anchor text into one or more tokens. (For example, see page 7, lines 6-10 and 18-20 and also, Fig. 4, steps 402 and 404.) The method includes determining a weight for the token. (For example, see page 7, line 27 - page 8, line 8 and also page 10, lines 28-31.) The method includes determining whether the weight assigned to the token exceeds a threshold token weight. (For example, see page 7, line 27 - page 8, line 8 and see also

page 11, lines 23-24). The method includes indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight. (For example, see page 7, line 27 - page 8, line 8).

Independent claim 9 recites a program product embedded in a computerized machine readable medium for including a document in an index in a hyperlinked environment. (For example, see Fig. 1, page 4, lines 24-34 and page 5, line 31 - page 6, line 2.) The program product includes instructions for receiving a document to be processed and locating a set of documents that include hyperlinks to the document. (For example, see page 6, lines 19-23). The program product includes instructions for retrieving anchortext associated with at least one of the hyperlinks and parsing the anchortext into one or more tokens. (For example, see page 7, lines 6-10 and 18-20 and also, Fig. 4, steps 402 and 404.) The program product includes instructions for determining a weight for the token. (For example, see page 7, line 27 - page 8, line 8 and also page 10, lines 28-31.) The program product further includes instructions for determining whether the weight assigned to the token exceeds a threshold token weight. (For example, see page 7, line 27 - page 8, line 8 and see also page 11, lines 23-24). The program product further includes instructions for indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight. (For example, see page 7, line 27 - page 8, line 8).

Independent claim 16 recites a computer system for indexing a document in hyperlinked environment. (For example, see Fig. 1, page 4, lines 24-34 and page 5, line 31 - page 6, line 2.) The computer system includes a receiver for receiving a document to be processed for inclusion in an index of documents a module for locating a set of

documents that include hyperlinks to the document. (For example, see Fig. 1 and page 6, lines 19-23). The computer system includes a module for retrieving anchor text associated with each hyperlink and a parsing module for parsing the anchor text into one or more tokens. (For example, see Fig. 1, page 7, lines 6-10 and 18-20 and also, Fig. 4, steps 402 and 404.) The computer system includes a module for determining a weight for the token. (For example, see page 7, line 27 - page 8, line 8 and also page 10, lines 28-31.) The module further determines whether the weight assigned to the token exceeds a threshold token weight. (For example, see page 7, line 27 - page 8, line 8 and see also page 11, lines 23-24). The module further indexes the document under the token, if the token weight assigned to the token exceeds the threshold token weight. (For example, see page 7, line 27 - page 8, line 8).

6. Grounds of Rejection to be Reviewed on Appeal

I. The Examiner mischaracterizes and misapplies the teachings of Bharat to the claimed invention.

II. Bharat, in combination with the Examiner's Official Notice, fail to teach or suggest all of the claimed limitations.

III. The Examiner's Official Notice in view of the shortcomings of Bharat is improper and contrary to the teachings of Bharat.

7. Argument

I. The Examiner mischaracterizes and misapplies the teachings of Bharat to the claimed invention.

Appellants have maintained during the prosecution and re-assert herein that the Examiner misapplies the teachings of Bharat and therefore the rejection of all pending claims as being obvious in view of Bharat is improper.

Bharat discusses a technique for ranking documents in hyperlink environment.

The Bharat system ranks the documents based on connectivity and content analysis. A detailed examination of the Specification indicates that Bharat uses a topic distillation technique by defining documents into nodes and defining weights of the nodes based on connectivity. More specifically, in the Bharat system, a hyperlinked document becomes a node (col. 2, lines 63-64), where documents “are Web pages connected to each other by hyperlinks.” (col. 2, lines 66-67). The nodes are initially scored based on connectivity to other nodes “and the number of terms that appear as unique sub-strings in the URL of the represented documents.” (col. 3, lines 11-13). The score for the document (node) “is a weighted sum of the number of directed edges to and from a node and the number of unique sub-strings of the URL that match a query term.”

Therefore, the Bharat system calculates a weight or a score using two different numbers. The first number is the number of connections to a node, see for example box 211 of Fig. 2. The second number is the number of sub-strings in the URL, which is short for the Universal Resource Locator, which is the same as a hypertext transfer protocol address, commonly recognized by the preface “http://.” For example, the Universal Resource Locator for the U.S. Patent Office is “http://www.uspto.gov.” There are many sub-pages within the parent domain of this URL, such as “http://www.uspto.gov/patents/main.htm” and “http://www.uspto.gov/main/-trademarks.htm”. Therefore, the strings of the URL are the text elements and sub-strings would be sub-combination of text elements contained in the URL.

The specific correlation of “sub-strings” to the query term of the Bharat system is emphasized by the scoring algorithm described in col. 5, lines 60-67. The value

“num_query_matches is the number of unique sub-strings of the URL of the page p that exactly match a term in the user’s query.” (col. 5, line63-65). Thus, the scoring and weighting mechanism of the Bharat system adds additional weighting when a search query finds more than one page having unique sub-string **in the URL** that matches the query term.

By contrast, the claimed invention of independent claims 1, 9 and 16 recite “retrieving **anchortext** associated with at least one of the hyperlinks” and “parsing the **anchortext** into one or more tokens.” The Bharat system is absolutely silent regarding anchortext. As defined in the specification and commonly recognized in the art, anchortext is text associated with a hyperlink. Fundamentally, anchortext **excludes** the URL itself, but rather represents viewable text that is presented to a user in lieu of the URL. In hypertext encoding, a hyperlink may be:

U.S. Patent and Trademark Office.

Through a browser application, this would generate the text “**U.S. Patent and Trademark Office**” being an active hyperlink (typically being shaded a blue color and underlined) that if selected, would cause the browser to then retrieve the HTML data from the URL address http://www.uspto.gov. In this example, “U.S. Patent Office” is the anchortext and “http://www.uspto.gov” is the URL. These are well known and expressly distinct terms.

On pages 2-3 of the final Office Action mailed July 11, 2006, the Examiner defines anchortext as being equivalent to “sub-strings in the URL.” (see, e.g. page 3, lines 1-2). As noted above, this is improper because sub-strings of the URL would be, as the term defines, sub-elements of larger strings of elements found **within the URL itself**

and anchor text is widely recognized as text that **accompanies** the active hyperlink and is not a sub-element or any element of the URL itself. At best, it is recognized that the anchor text provides a description of the underlying referenced document or location designate by the URL (e.g. anchor text of “U.S. Patent and Trademark Office” describes the URL “http://www.uspto.gov.”)

In the “Response to Arguments” section beginning on page 7 of the Office Action dated February 8, 2006, the Examiner asserts the position that it is reasonable to interpret the term “anchor text” as including the URL, to which Appellants must respectfully disagree. The Examiner makes this generalized interpretation on the idea that a web-page author has the discretion to use any text for the anchor text and the anchor text can be the same of the URL. This would be the example where an author encodes a hyperlink of www.uspto.gov. The Examiner’s interpretation is fundamentally flawed in this example, the URL does **not** become the anchor text, but rather the URL merely shares the same characters as the anchor text. In this case, the URL is the first “www.uspto.gov” instance (i.e. and the anchor text is second “www.uspto.gov” instance. While sharing the same characters, the URL and the anchor text are treated separately and distinctly in both the Bharat system and the claimed invention. The Bharat system performs weighting calculations based on the first instance (URL) and is completely silent regarding the treatment of the second instance (anchor text). By contrast, the claimed invention claims functions associated with the second instance (anchor text) and not the first instance (URL). Therefore, the Examiner improperly interprets the term anchor text to include a URL and then improperly applies this improper interpretation to the presently claimed invention.

Therefore, Appellant respectfully submits that the rejection is improper because the Examiner mischaracterizes and misapplies the teachings of Bharat.

II. Bharat, in combination with the Examiner's Official Notice, fail to teach or suggest all of the claimed limitations.

As described above, Appellant maintains that the Examiner misapplies the concept of URL sub-strings to the claimed "anchortext" of the present invention. Additionally, the Examiner's interpretation of Bharat produces a system different than the claimed invention.

First, by way of terminology, the Examiner determined the claimed anchortext is equivalent to the sub-strings in the URL of Bharat. Claim 1 recites "parsing the anchortext into one or more tokens." Therefore, by correlation, a token of the Bharat system is a parsed element of the anchortext (e.g. sub-strings of the URL). By the Examiner's interpretation, the token is a sub-string of the sub-string of the URL, or a sub-sub-string.

In the Office Action finally rejecting all pending claims, the Examiner indicates that Bharat discusses determining a weight for the token and determining whether the weight exceeds a threshold amount on page 3, lines 3-9. Appellants submit this improper because, by definition, Bharat performs a weighting function on the URL and not the anchortext.

As claimed herein, the tokens are parsed elements of the anchortext. The tokens are not and cannot be interpreted to be sub-sub-strings of the URL for the reasons stated above. Therefore, the Examiner improperly asserts Bharat as teaching or suggesting determining a weight for the token because the claimed token is a parsed element of the

anchortext and not a sub-sub-string of the URL. Additionally, the Examiner takes official notice regarding “indexing” and therefore the Examiner’s official notice fails to cure this explicit deficiency of Bharat.

III. The Examiner’s Official Notice in view of the shortcomings of Bharat is improper and contrary to the teachings of Bharat.

In regards to the presently claimed element of “indexing the document under the token,” the Examiner takes official notice that it would have been obvious to include indexing techniques with the system of Bharat. Appellants maintain that this improper in view of the teachings of Bharat with respect to claimed invention.

Bharat discusses a system that indexes the nodes (which represent documents) based on two factors: (1) number of connections to the nodes; and (2) the number unique sub-strings **in the URL**. As described above, under the Examiner’s interpretive application of Bharat to the claimed invention, the node (e.g. document, a web page) is indexed based on a sub-sub-string **of the URL**. This is in direct contradiction to the teachings of Bharat because Bharat uses the two factors noted above. This is also in direct contradiction to the claimed invention because the claimed documents are indexed under the token (which is a parsed element **of the anchortext**).

Therefore, Appellants submit that the Examiner’s Official Notice is improper because it is contrary to the teachings of Bharat. Appellants also submit that the even with the Examiner’s Official Notice, the rejection is improper because the combination of Bharat and Examiner’s Official Notice still fails to teach or suggest all of the claimed elements recited herein because Bharat teaches functionality based on the URL itself and not the anchortext as claimed herein.

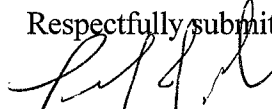
8. Conclusions

For the reasons set forth above, Appellants respectfully request that the Board reverse the final rejections of the claims by the Examiner under 35 U.S.C. § 103(a) and indicate that claims 1-12 and 13-14 are allowable.

Dated: May 16, 2007

THIS CORRESPONDENCE IS BEING SUBMITTED
ELECTRONICALLY THROUGH THE PATENT AND
TRADEMARK OFFICE EFS FILING SYSTEM ON
May 16, 2007.

Respectfully submitted,



Timothy J. Bechen

Reg. No. 48,126

DREIER LLP

499 Park Ave.

New York, New York 10022

Tel : (212) 328-6100

Fax: (212) 328-6101

Customer No. 61834

Claims Appendix

1. (Original) A method for including a document in an index in a hyperlinked environment, comprising the acts of:
 - receiving a document to be processed;
 - locating a set of documents that include hyperlinks to the document;
 - retrieving anchortext associated with at least one of the hyperlinks;
 - parsing the anchortext into one or more tokens;
 - for each token:
 - determining a weight for the token,
 - determining whether the weight assigned to the token exceeds a threshold token weight; and
 - indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight.
2. (Original) The method of Claim 1, wherein the indexing act comprises including in the index an indication of weight for each token under which each page is indexed.
3. (Original) The method of Claim 1, wherein the indexing act comprises assigning to the token a location within the index corresponding to part of the page being indexed that is allocated for tokens having a higher degree of importance than other tokens in the same page.
4. (Original) The method of Claim 3, wherein the indexing act comprises assigning to the token a location within the index that corresponds to the beginning of the page being indexed.

5. (Original) The method of Claim 1, wherein the weight of each token is based on its frequency of occurrence within the index.
6. (Original) The method of 1, wherein the act of determining a weight comprises: determining a first frequency at which the anchortext appears in the index; determining a second frequency at which each token derived from the anchortext appears in the index; and assigning a weight to the token, wherein the weight is a function of the first and second frequencies.
7. (Original) The method of Claim 6, further comprising dividing the first frequency by the second frequency to produce a weight quotient; and multiplying the weight quotient by an anchor text count for the token.
8. (Original) The method of 6 further comprising determining a normalized weight for each token.
9. (Previously Presented) A program product embedded in a computerized machine-readable medium for including a document in an index in a hyperlinked environment, comprising the instructions for:
 - receiving a document to be processed;
 - locating a set of documents that include hyperlinks to the document;
 - retrieving anchortext associated with each hyperlink;
 - parsing the anchortext into one or more tokens;
 - and program instructions for each token comprising instructions for:
 - determining a weight for the token,
 - determining whether the weight assigned to the token exceeds a threshold token weight; and

indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight.

10. (Original) The computer program product of Claim 9 wherein the indexing instruction comprises including in the index an indication of weight for each token under which each page is indexed.

11. (Original) The computer program product of Claim 9, wherein the weight of each token is based on its frequency of occurrence within the index.

12. (Original) The computer program product of Claim 9, wherein the indexing act comprises assigning to the token a location within the index that corresponds to the beginning of the page being indexed.

13. (Cancelled)

14. (Original) The computer program product of Claim 9, wherein the instruction of determining a weight comprises:

determining a first frequency at which the anchortext appears in the index;

determining a second frequency at which each token derived from the anchortext appears in the index; and

assigning a weight to the token, wherein the weight is a function of the first and second frequencies.

15. (Previously Presented) The program product of claim 9, further comprising the instruction of determining a normalized weight for each token.

16. (Previously Presented) A computer system for indexing a document in a hyperlinked environment, comprising:

a receiver for receiving a document to be processed for inclusion in an index of documents;

a module for locating a set of documents that include hyperlinks to the document;

a module for retrieving anchortext associated with each hyperlink;

a parsing module for parsing the anchortext into one or more tokens;

a module for:

determining a weight for the token,

determining whether the weight assigned to the token exceeds a threshold token weight; and

indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight.

Evidence Appendix

No evidence has been submitted or relied upon in the instant appeal.

Related Proceedings Appendix

There are no related proceedings which are related to or would have a bearing on the instant appeal.